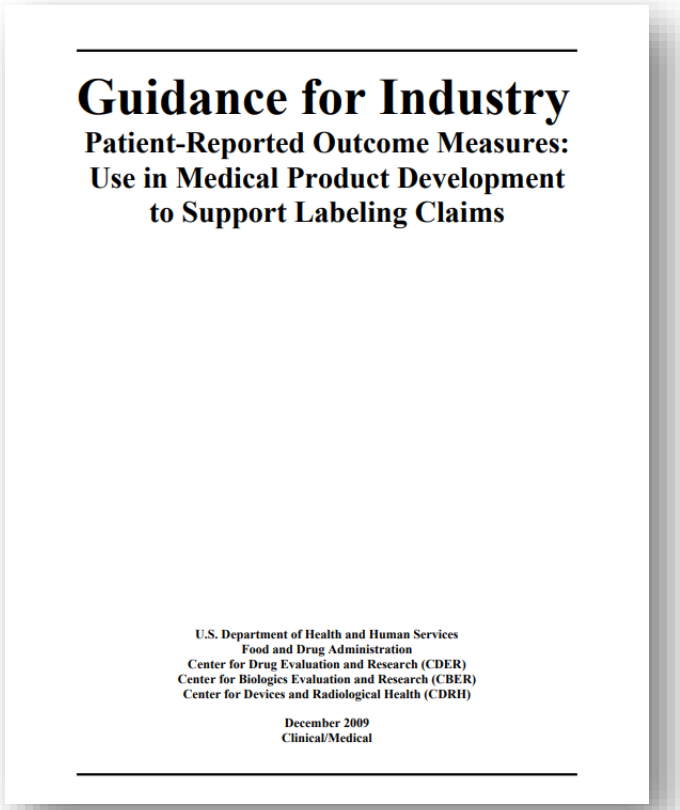
Enabling Patient-Reported Outcome Measures (PROMs) in Clinical Trials, Exemplified by Cardiovascular Trials

Friday, September 24, 2021, 1:00-2:00 p.m. ET

Theresa Coles, PhD
Assistant Professor
Department of Population Health Sciences
Duke University School of Medicine

Kevin Weinfurt, PhD
Professor and Vice Chair of Research
Department of Population Health Sciences
Duke University School of Medicine

Patient-Reported Outcomes (PROs)



Guidance for Industry

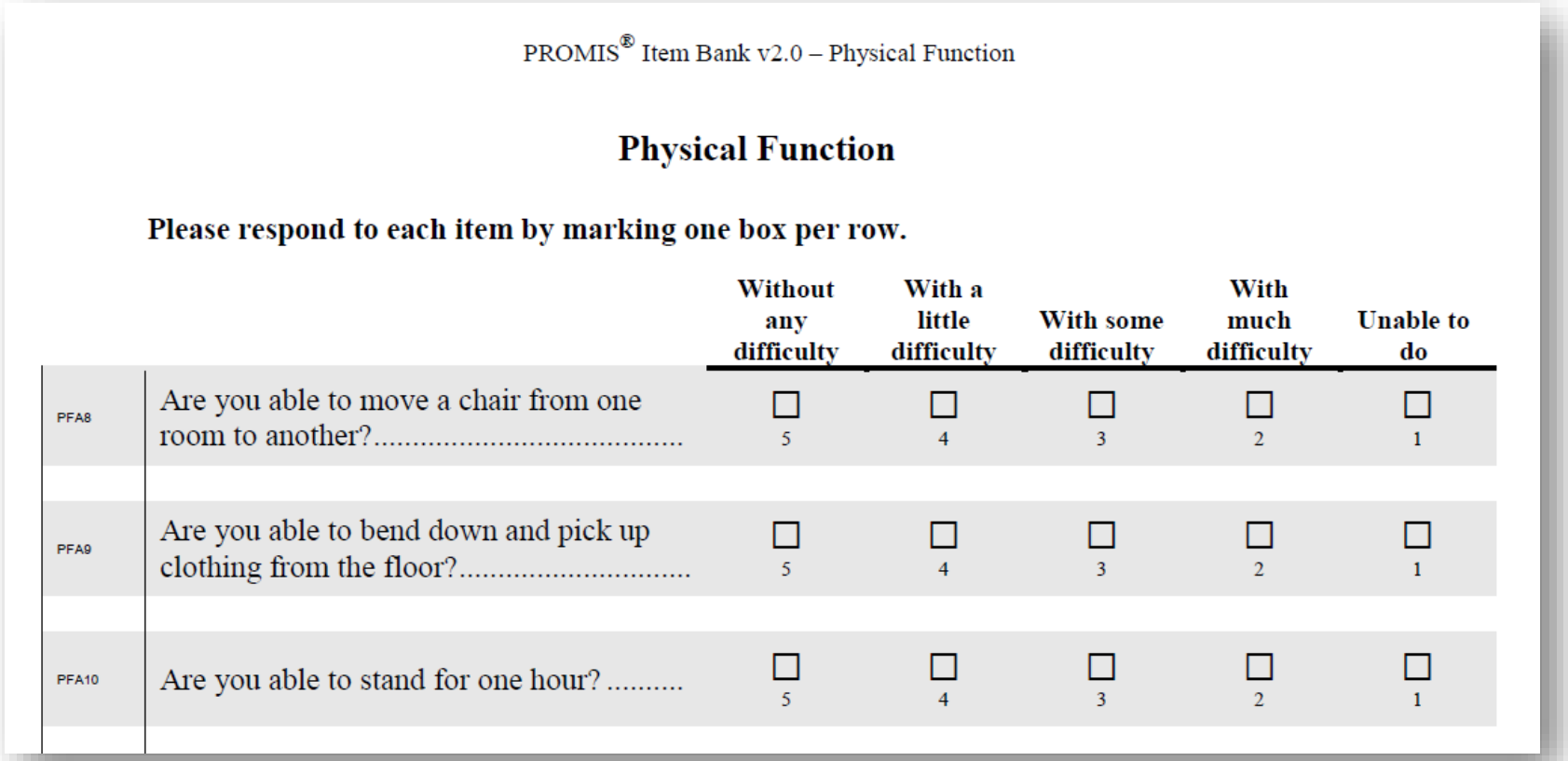
Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

December 2009
Clinical/Medical

"A **PRO** is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else."

Patient-Reported Outcome Measures (PROMs)



PROMIS® Item Bank v2.0 – Physical Function

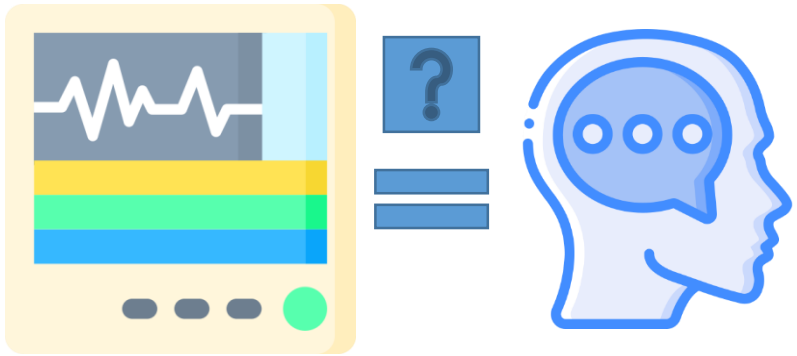
Physical Function

Please respond to each item by marking one box per row.

		Without any difficulty	With a little difficulty	With some difficulty	With much difficulty	Unable to do
PFA8	Are you able to move a chair from one room to another?.....	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFA9	Are you able to bend down and pick up clothing from the floor?.....	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1
PFA10	Are you able to stand for one hour?	☐ 5	☐ 4	☐ 3	☐ 2	☐ 1

PROMs are the measurement tool – a questionnaire

Value of PROMs



Peak V02

r = ?

Flynn et al, 2009

Peak V02

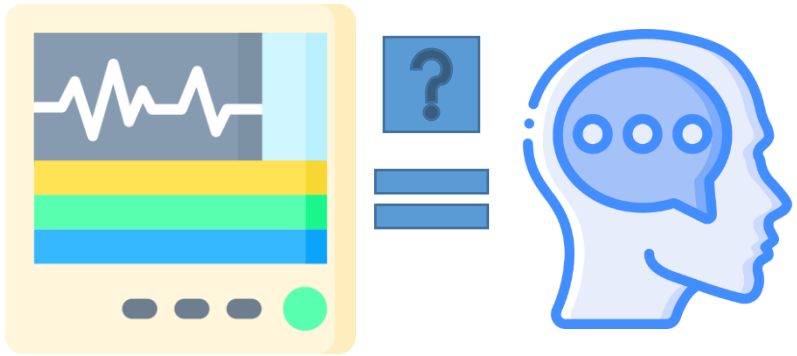
r = .3

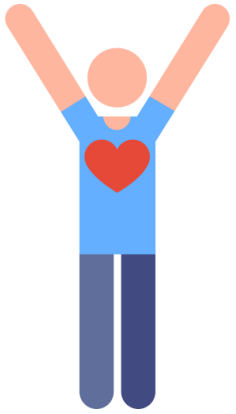





Flynn et al, 2009

Value of PROMs





PROMs in Explanatory and Pragmatic Trials

Kevin Weinfurt

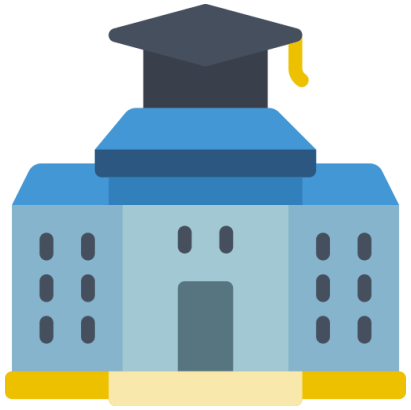
PROM Use for Endpoints in Trials

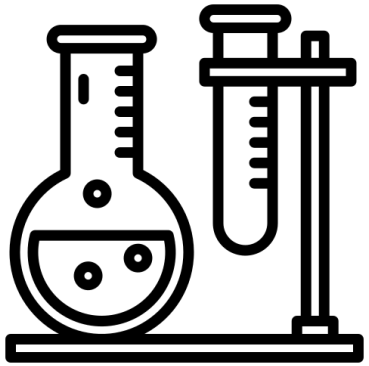
Traditional Explanatory Trials

Pragmatic Trials

- Use limited if PROM not collected as part of usual clinical care
- Increasing use of PROMs in clinical care → inclusion in electronic health record (EHR)
- PROMs used for endpoints in NIH *Pragmatic and Implementation Studies for the Management of Pain to Reduce Opioid Prescribing* (PRISM)

Think Tank





47 stakeholder participants

FDA

Coles et al. Health Qual Life Outcomes (2021) 19:164
https://doi.org/10.1186/s12955-021-01800-1

Health and Quality of Life Outcomes

COMMENTARY Open Access

Enabling patient-reported outcome measures in clinical trials, exemplified by cardiovascular trials

Theresa M. Coles[1*], Adrian F. Hernandez[1], Bryce B. Reeve[1], Karon Cook[2], Michael C. Edwards[3,4], Marc Boutin[5], Elizabeth Bush[6], Arnold Degboe[7], Lothar Roessig[8], Amy Rudolph[9], Pauline McNulty[10], Nikunj Patel[7], Trish Kay-Mugford[9], Margaret Vernon[11], Michael Woloschak[12], Gustavo Buchele[13], John A. Spertus[14,15], Matthew T. Roe[1], Denise Bury[16] and Kevin Weinfurt[1]

In our talk today, we will cover 3 things…

1. What are the current challenges for integrating PROMs in clinical trials?
2. What are we proposing to do?
3. What is left to do?

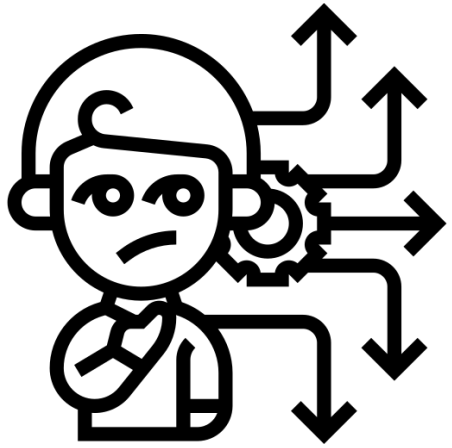
Challenges with integrating PROs in clinical trials

Culture

Budget and Time

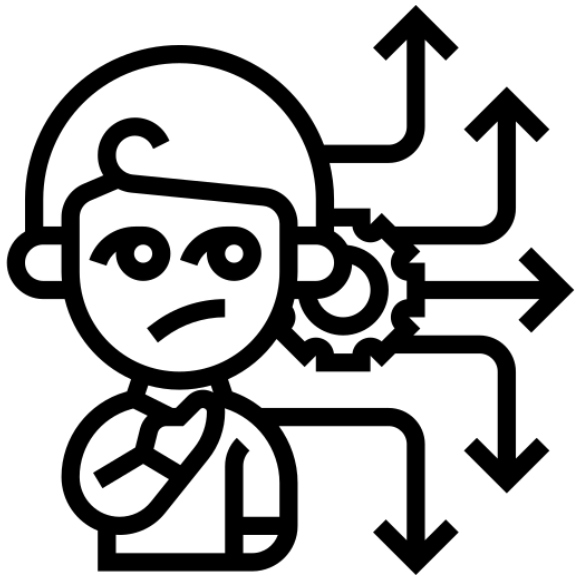
Unclear interpretation of scores

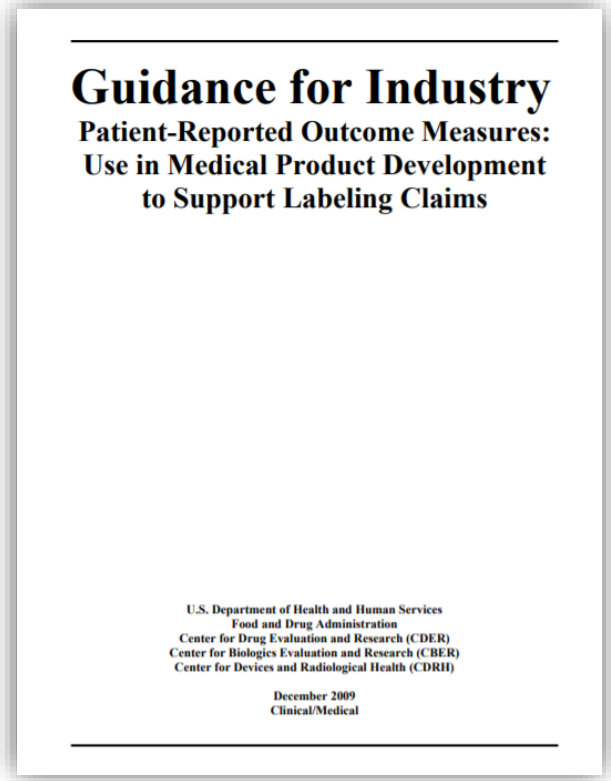
Missing information



Uncertainty

What are the evidentiary requirements for a PROM in the clinical trial we are starting in a few months?



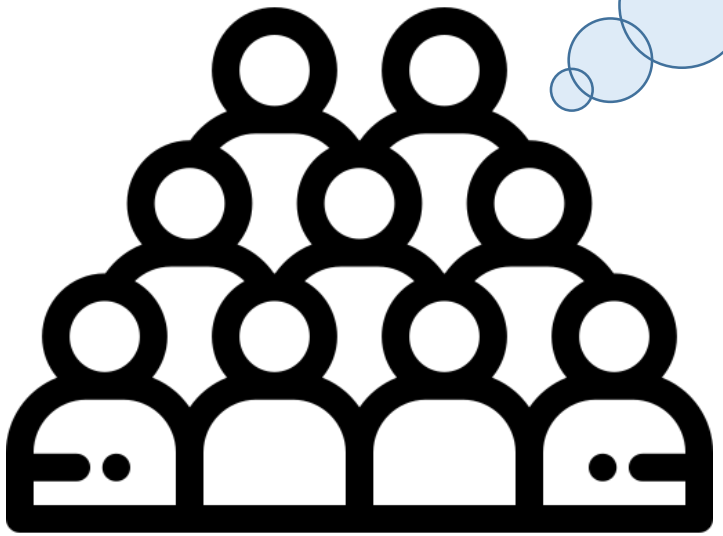


Guidance for Industry

Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

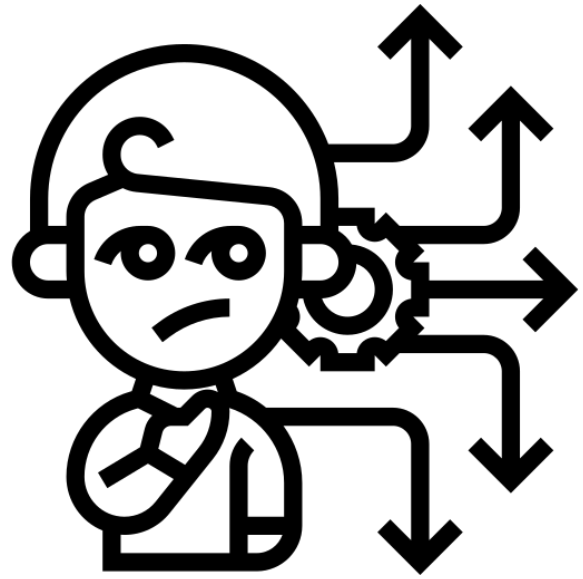
December 2009
Clinical/Medical



What are the evidentiary requirements for a PROM in the clinical trial we are starting in a few months?

Answer:

Modern validity theory may hold the solution.

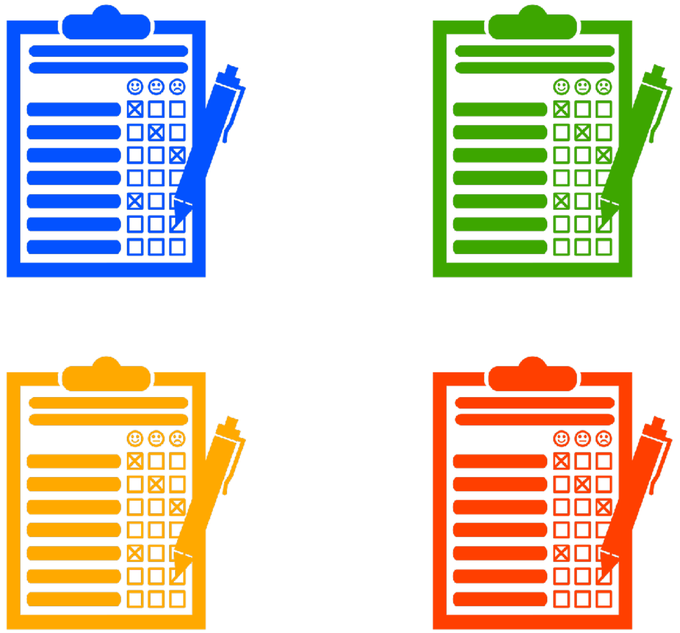


Department of Population Health Sciences
Duke University School of Medicine

Modern Validity Theory

Kevin Weinfurt

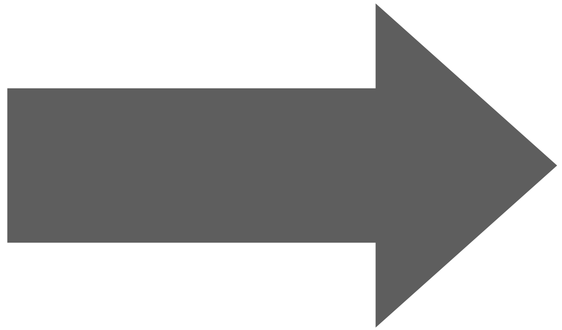
Health Measures



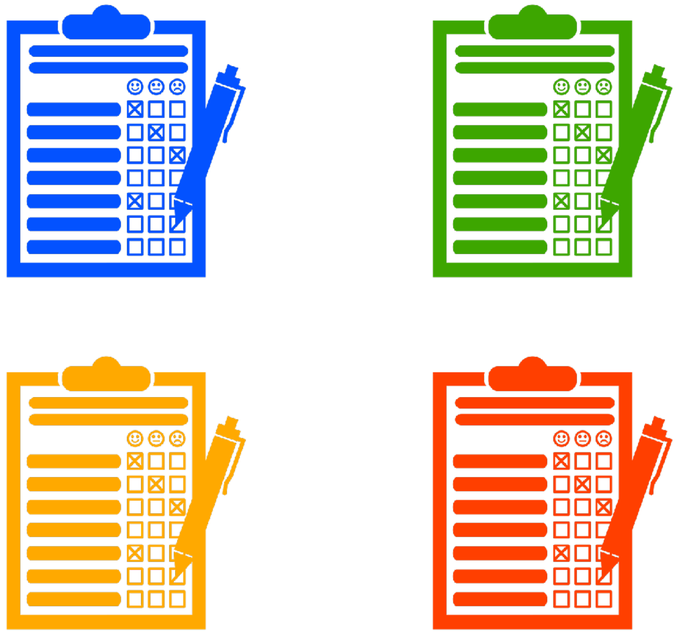
Frameworks for Evaluating Measures

Psychology

Educational Testing



Health Measures



Validation work as “stamp collecting”

Bruno Zumbo, 2014

Content Validity Study 1

Discriminant Validity Study

Structural Validity Study

Measurement Invariance Study 2

Test-retest Reliability Study

Known Groups Validity Study 1

Internal Consistency Reliability Study

Content Validity Study 2

Convergent Validity Study 1

Responsiveness to Change Study

Known Groups Validity Study 2

Predictive Validity Study 1

Measurement Invariance Study 1

Convergent Validity Study 2

Cognitive Interview Study

Content Validity Study 1

Discriminant Validity Study

Structural Validity

Measurement Invariance Study 2

Test-retest Reliability Study

Known Groups Validity Study 1

Internal Consistency Reliability Study

Convergent Validity Study 1

Responsiveness to Change Study

Known Groups Validity Study 2

Predictive Validity Study 1

Measurement Invariance Study 1

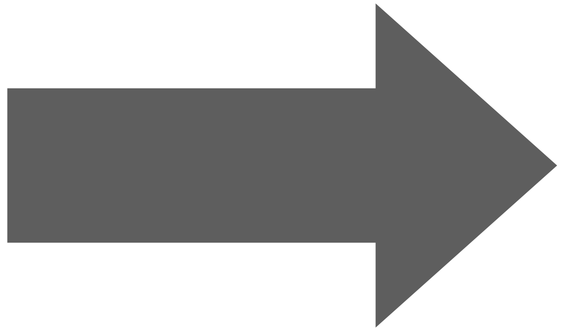
Convergent Validity Study 2

Cognitive Interview Study

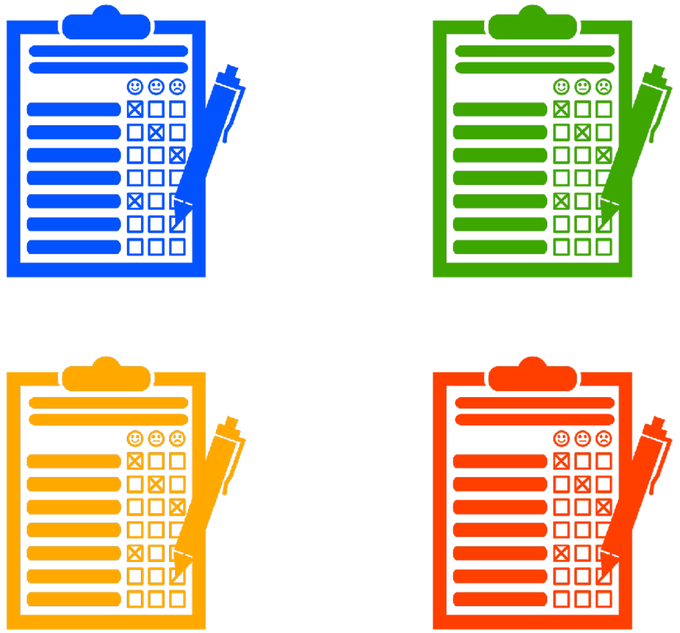
Frameworks for Evaluating Measures

Psychology

Educational Testing



Health Measures

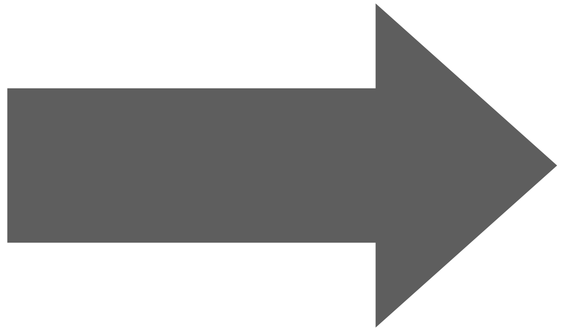


Modern

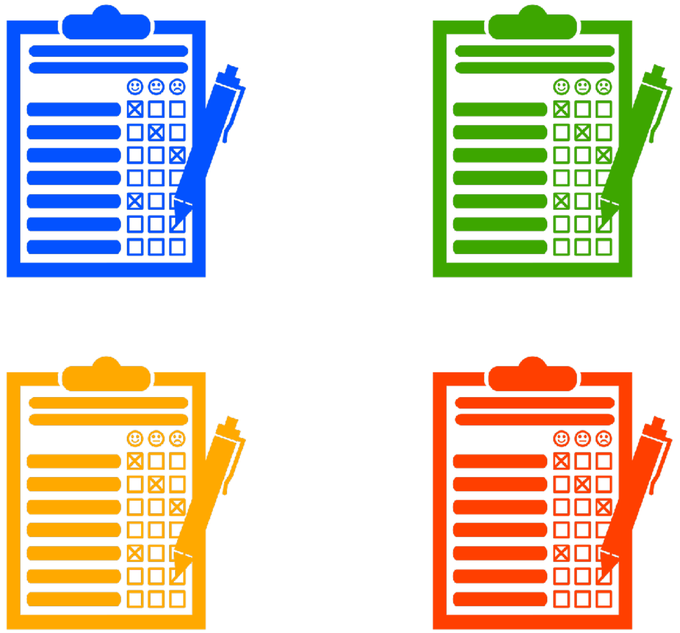
Frameworks for Evaluating Measures

Psychology

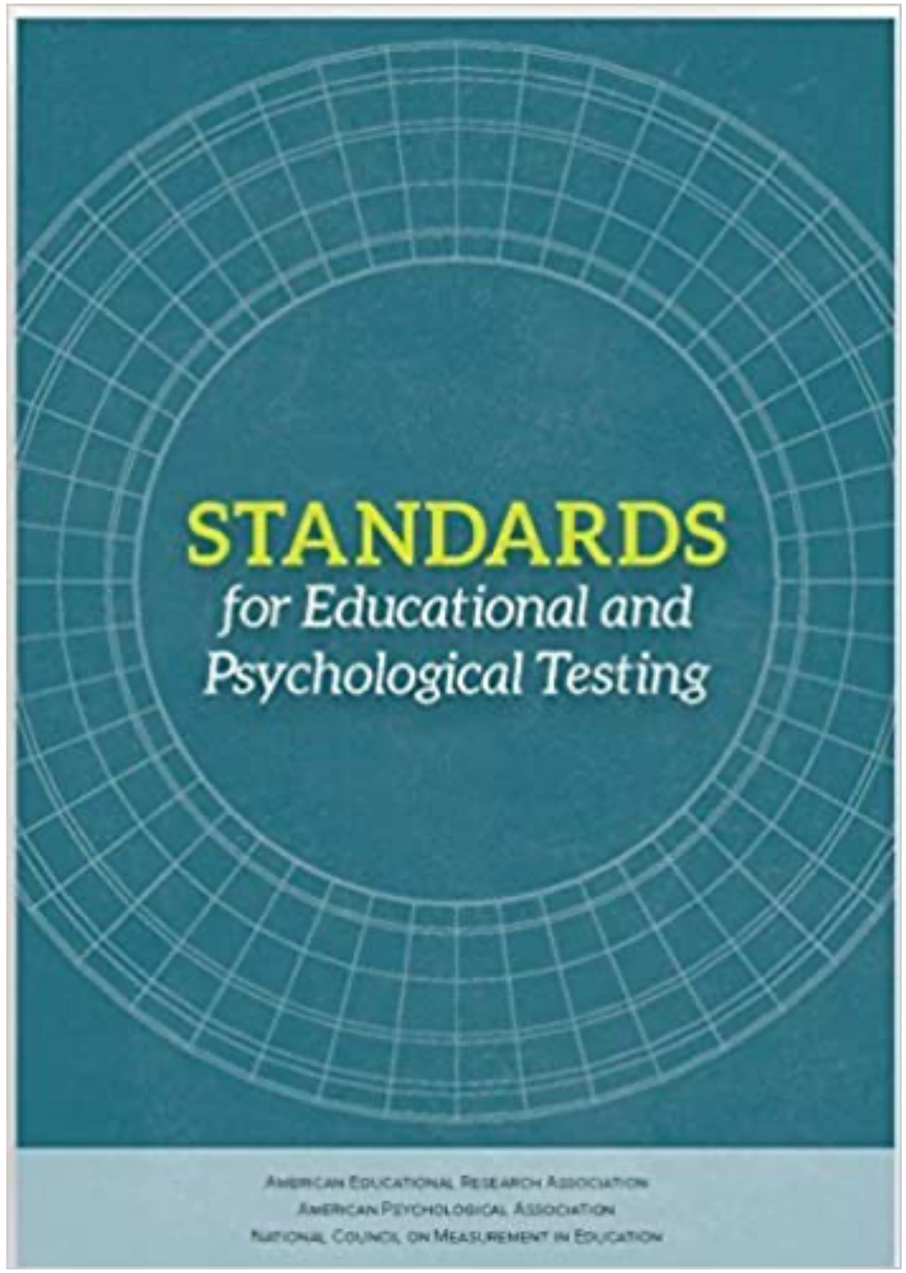
Educational Testing



Health Measures



Modern Validity Theory actually includes many different (sometimes conflicting) perspectives and approaches



Developed jointly by:

- American Educational Research Association (AERA)
- American Psychological Association (APA)
- National Council on Measurement in Education (NCME)

Reflects accumulated experience and insight into testing in high-stakes situations for over 50 years (first edition was 1966)

Current edition is 2014

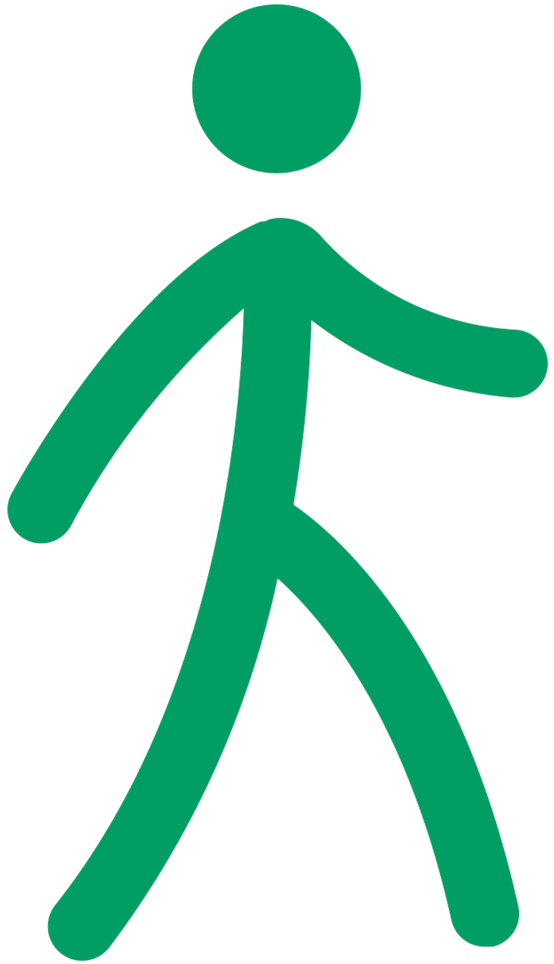
Validity

"…the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11, *2014 Standards for Educational and Psychological Testing*)

Validity

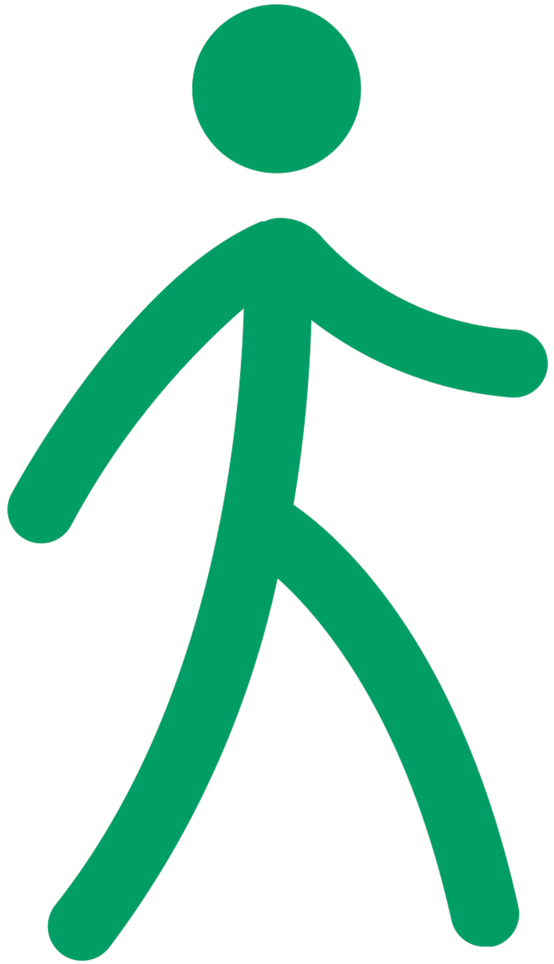
"…the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11, *2014 Standards for Educational and Psychological Testing*)

6 Minute Walk Test



To what degree do evidence and theory support its interpretation as a measure of…

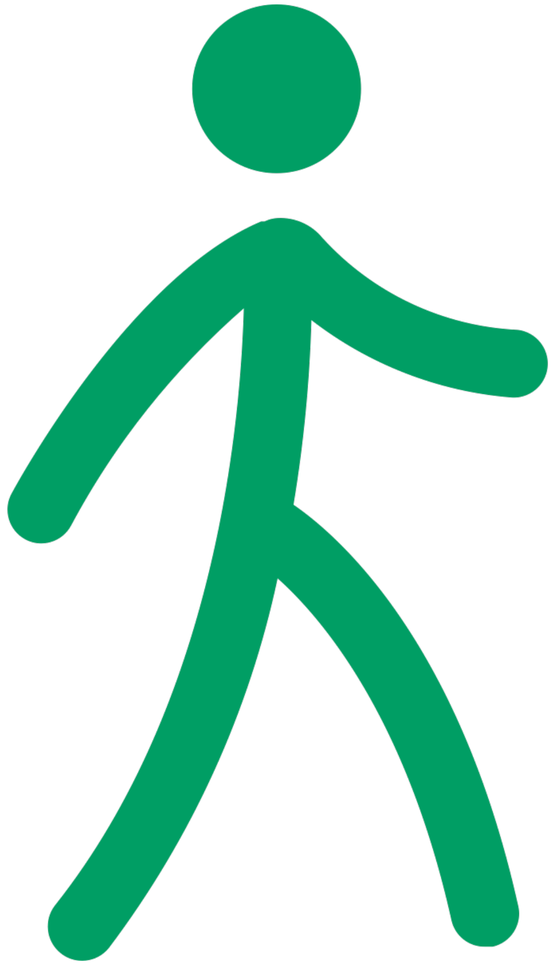
6 Minute Walk Test



To what degree do evidence and theory support its interpretation as a measure of…

How Far You Can Walk in 6 Minutes?

6 Minute Walk Test

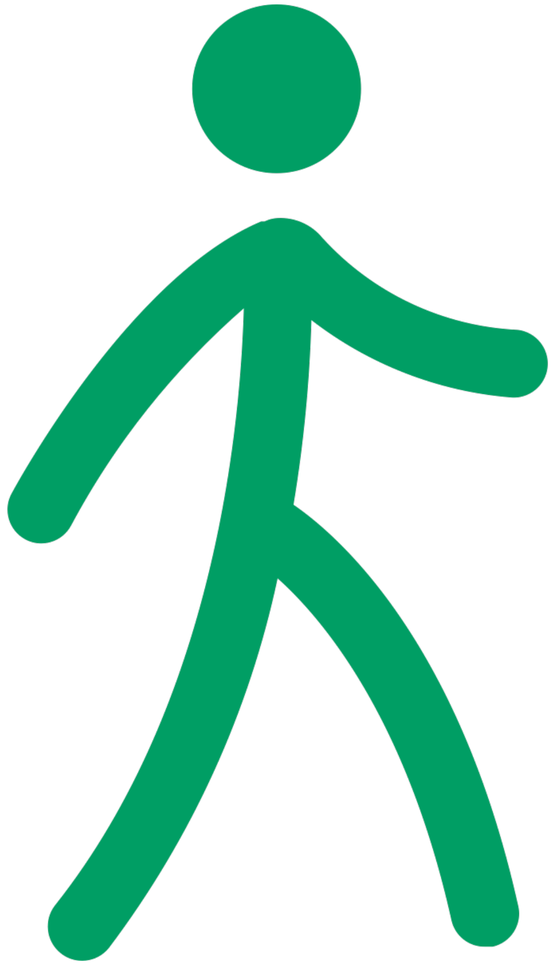


To what degree do evidence and theory support its interpretation as a measure of…

How Far You Can Walk in 6 Minutes?

Mobility?

6 Minute Walk Test



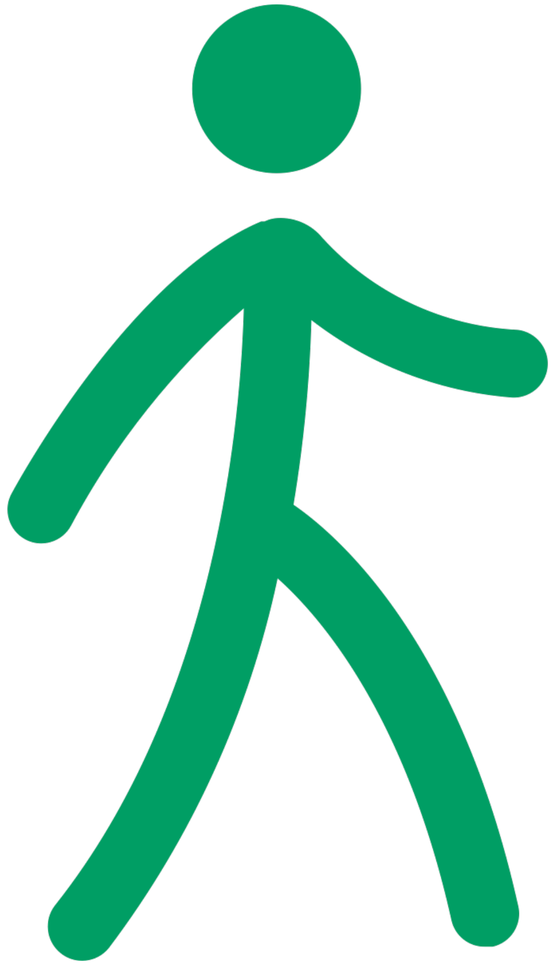
To what degree do evidence and theory support its interpretation as a measure of…

How Far You Can Walk in 6 Minutes?

Mobility?

Aerobic capacity?

6 Minute Walk Test



To what degree do evidence and theory support its interpretation as a measure of…

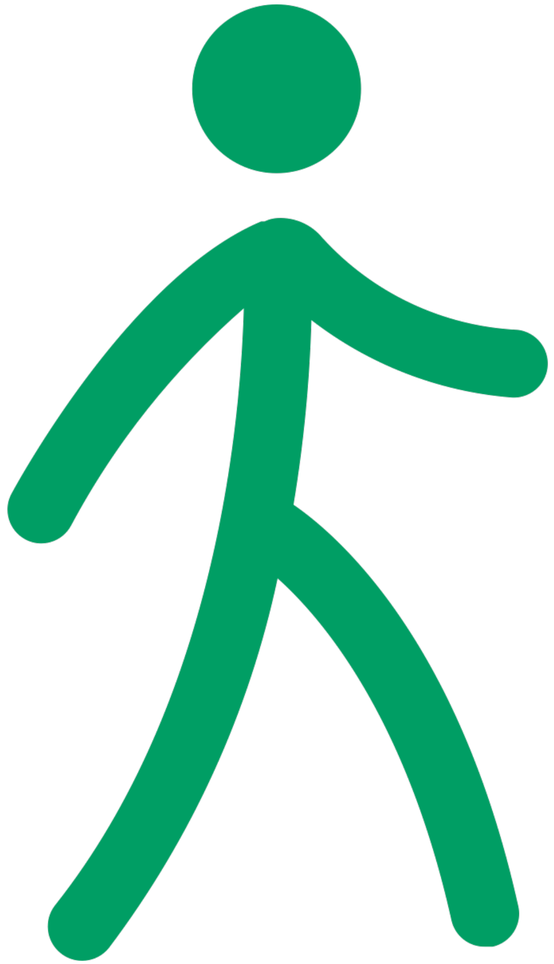
How Far You Can Walk in 6 Minutes?

Mobility?

Aerobic capacity?

Endurance?

6 Minute Walk Test



To what degree do evidence and theory support its interpretation as a measure of…

How Far You Can Walk in 6 Minutes?

Mobility?

Aerobic capacity?

Endurance?

Physical Functioning?

Validity

"…the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11, *2014 Standards for Educational and Psychological Testing*)

Validity

"…the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11, *2014 Standards for Educational and Psychological Testing*)

FDA's *Fit-for-purpose* = "A conclusion that the level of validation associated with a biomarker or [clinical outcome assessment] is sufficient to support its proposed use." (BEST Glossary)

Validity

"…the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11, *2014 Standards for Educational and Psychological Testing*)

FDA's *Fit-for-purpose* = "A conclusion that the level of validation associated with a biomarker or [clinical outcome assessment] is sufficient to support its proposed use." (BEST Glossary)

Fit-for-purpose = Valid

Validity

"…the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11, *2014 Standards for Educational and Psychological Testing*)

FDA's *Fit-for-purpose* = "A conclusion that the level of validation associated with a biomarker or [clinical outcome assessment] is sufficient to support its proposed use." (BEST Glossary)

Fit-for-purpose = Valid

Validation

2014 Standards for Educational and Psychological Testing

Validation

"…can be viewed as a process of constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed use" (p. 11).

2014 Standards for Educational and Psychological Testing

Validation

"…can be viewed as a process of constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed use" (p. 11).

2014 Standards for Educational and Psychological Testing

Validation Under the Argument-Based Approach

Validation Under the Argument-Based Approach

1. State how you intend to interpret and use scores from the measure

Validation Under the Argument-Based Approach

1. State how you intend to interpret and use scores from the measure
2. Identify the assumptions and threats to the assumptions underlying the proposed interpretation/use of scores

Validation Under the Argument-Based Approach

1. State how you intend to interpret and use scores from the measure
2. Identify the assumptions and threats to the assumptions underlying the proposed interpretation/use of scores
3. Collect or cite supporting evidence for each assumption

Validation Under the Argument-Based Approach

1. State how you intend to interpret and use scores from the measure
2. Identify the assumptions and threats to the assumptions underlying the proposed interpretation/use of scores
3. Collect or cite supporting evidence for each assumption
4. Evaluate . . .
    A. Comprehensiveness and coherence of the rationale
    B. Strength of support for the assumptions

Validation Under the Argument-Based Approach

1. **State how you intend to interpret and use scores from the measure**
2. Identify the assumptions and threats to the assumptions underlying the proposed interpretation/use of scores
3. Collect or cite supporting evidence for each assumption
4. Evaluate . . .
    A. Comprehensiveness and coherence of the rationale
    B. Strength of support for the assumptions

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

Adapted from 2014 Standards for Educational and Psychological Testing

Validation Under the Argument-Based Approach

1. State how you intend to interpret and use scores from the measure
2. **Identify the assumptions and threats to the assumptions underlying the proposed interpretation/use of scores**
3. Collect or cite supporting evidence for each assumption
4. Evaluate . . .
    A. Comprehensiveness and coherence of the rationale
    B. Strength of support for the assumptions

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course

2. Content domain of the MAT is consistent with these prerequisite skills

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course

2. Content domain of the MAT is consistent with these prerequisite skills

3. Test scores can be generalized across relevant sets of items (i.e., MAT has many possible items and each student is given a subset of them; so assumption is that scores are relatively consistent regardless of which set of MAT items you receive)

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course

2. Content domain of the MAT is consistent with these prerequisite skills

3. Test scores can be generalized across relevant sets of items (i.e., MAT has many possible items and each student is given a subset of them; so assumption is that scores are relatively consistent regardless of which set of MAT items you receive)

4. Test scores are not unduly influenced by ancillary variables, such as writing ability

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course

2. Content domain of the MAT is consistent with these prerequisite skills

3. Test scores can be generalized across relevant sets of items (i.e., MAT has many possible items and each student is given a subset of them; so assumption is that scores are relatively consistent regardless of which set of MAT items you receive)

4. Test scores are not unduly influenced by ancillary variables, such as writing ability

5. Success in the advanced course can be validly assessed

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course

2. Content domain of the MAT is consistent with these prerequisite skills

3. Test scores can be generalized across relevant sets of items (i.e., MAT has many possible items and each student is given a subset of them; so assumption is that scores are relatively consistent regardless of which set of MAT items you receive)

4. Test scores are not unduly influenced by ancillary variables, such as writing ability

5. Success in the advanced course can be validly assessed

6. Test takers with high scores on the test will be more successful in the advanced course than test takers with low scores on the test

Adapted from 2014 Standards for Educational and Psychological Testing

How do we know the type and amount of evidence needed in any given situation?

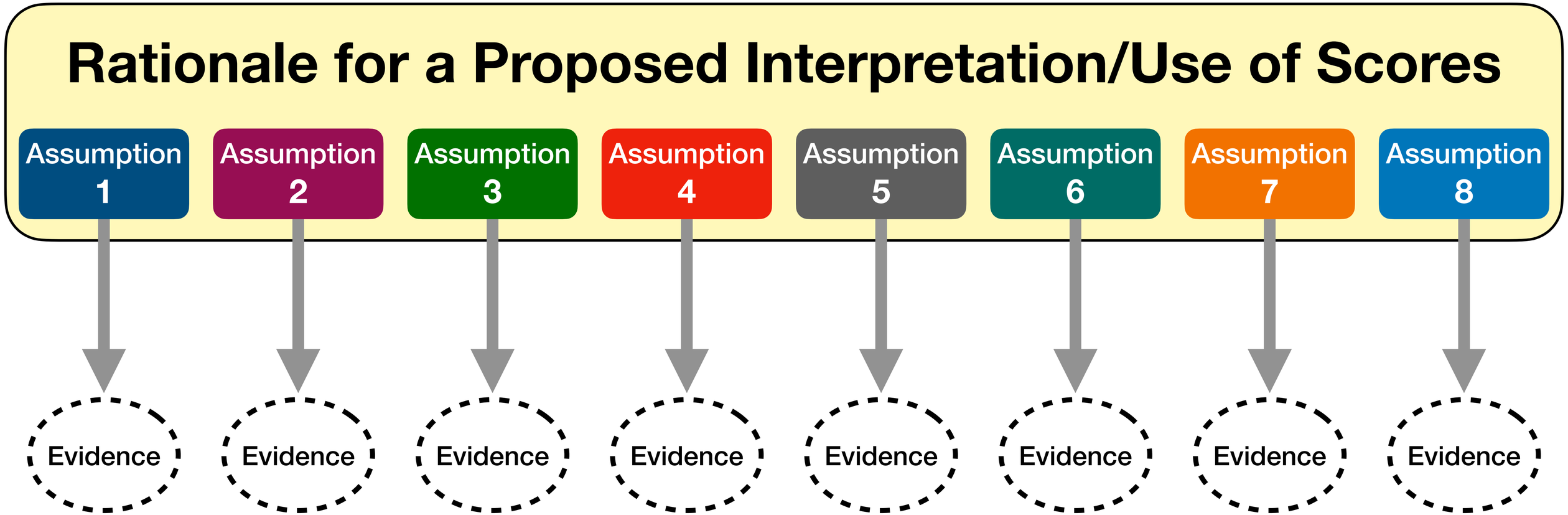
How do we know the type and amount of evidence needed in any given situation?

Each assumption of the rationale highlights a need for evidence.

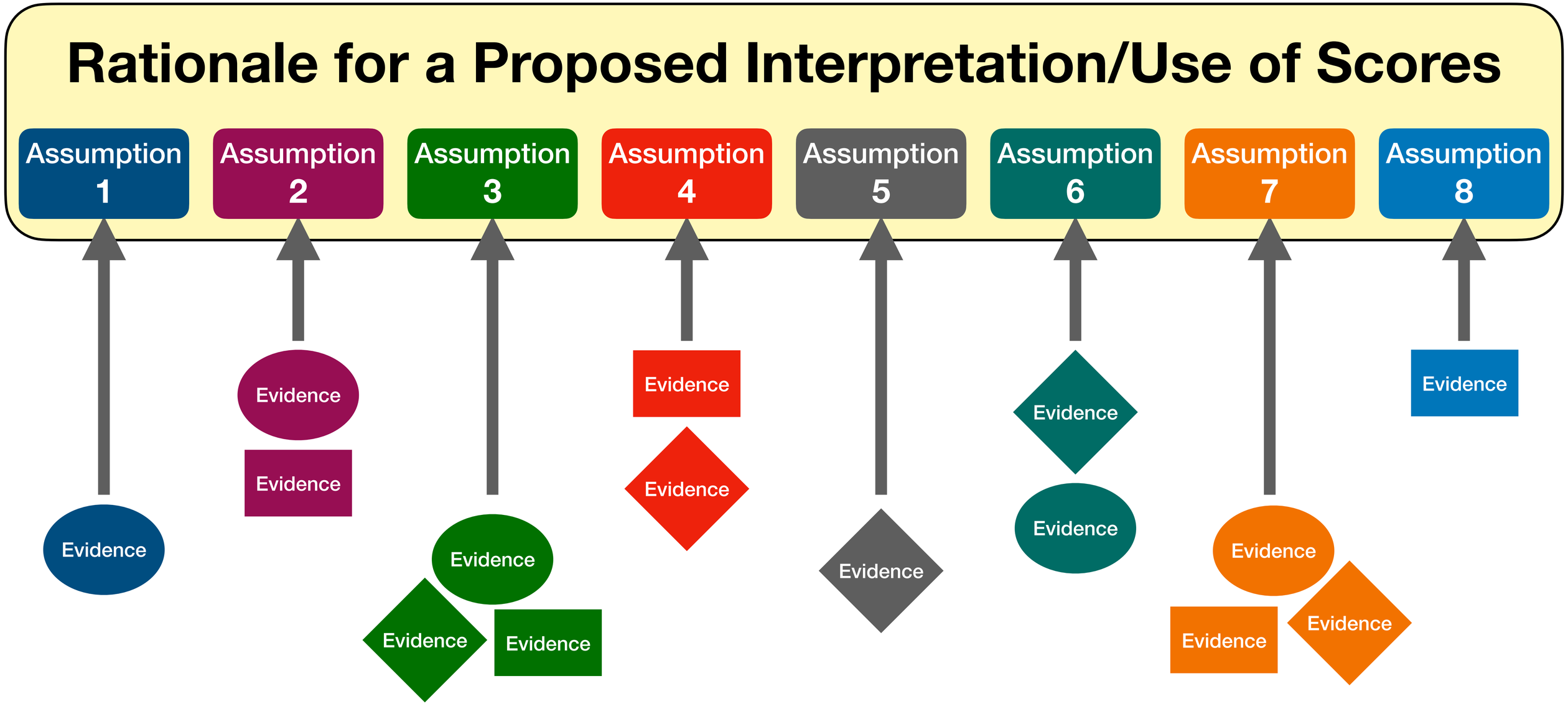
Rationale for a Proposed Interpretation/Use of Scores

- Assumption 1
- Assumption 2
- Assumption 3
- Assumption 4
- Assumption 5
- Assumption 6
- Assumption 7
- Assumption 8

Each assumption of the rationale highlights a need for evidence.



Multiple types and sources of evidence might be used to evaluate each assumption.



Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course

2. Content domain of the MAT is consistent with these prerequisite skills

3. Test scores can be generalized across relevant sets of items (i.e., MAT has many possible items and each student is given a subset of them; so assumption is that scores are relatively consistent regardless of which set of MAT items you receive)

4. Test scores are not unduly influenced by ancillary variables, such as writing ability

5. Success in the advanced course can be validly assessed

6. Test takers with high scores on the test will be more successful in the advanced course than test takers with low scores on the test

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

1. Certain mathematical skills are prerequisite for the advanced course
2. Content domain of the MAT is consistent with these prerequisite skills
3. Test scores can be generalized across relevant sets of items (i.e., MAT has many possible items and each student is given a subset of them; so assumption is that scores are relatively consistent regardless of which set of MAT items you receive)
4. Test scores are not unduly influenced by ancillary variables, such as writing ability
5. Success in the advanced course can be validly assessed
6. Test takers with high scores on the test will be more successful in the advanced course than test takers with low scores on the test

Adapted from 2014 Standards for Educational and Psychological Testing

Scores on the Mathematics Achievement Test (MAT) can be used to assess readiness for an advanced course.

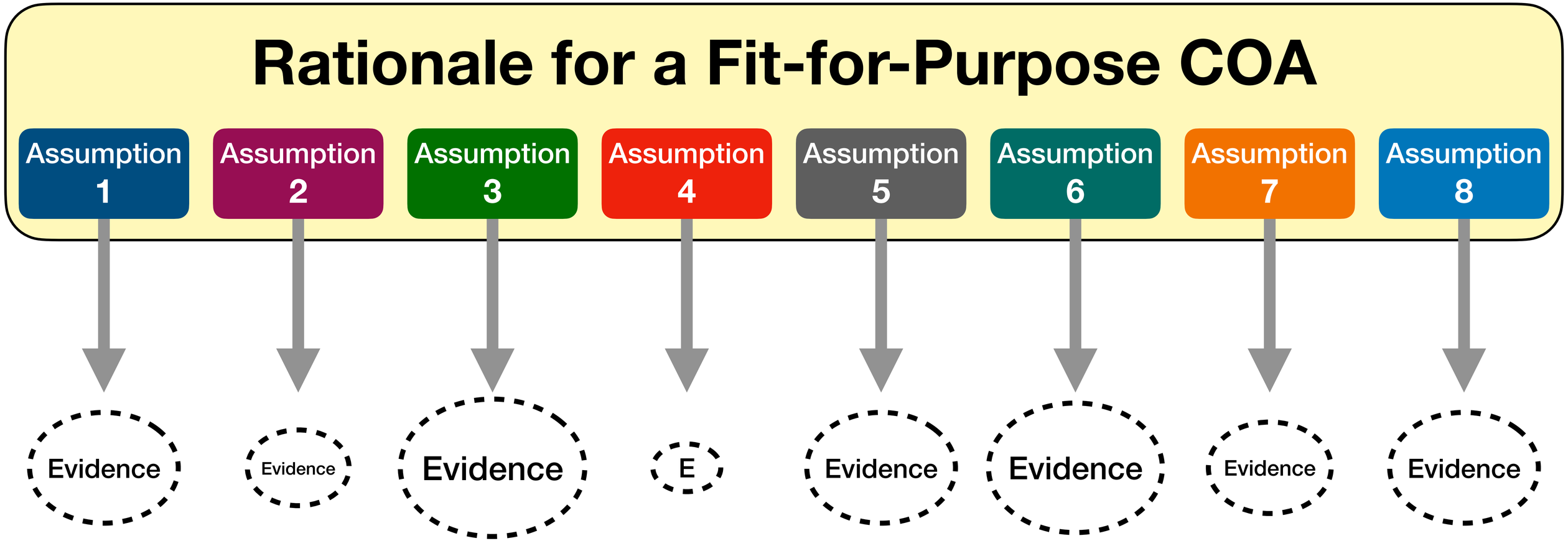
1. Certain mathematical skills are prerequisite for the advanced course
2. Content domain of the MAT is consistent with these prerequisite skills
3. Test scores can be generalized across relevant sets of items (i.e., MAT has many possible items and each student is given a subset of them; so assumption is that scores are relatively consistent regardless of which set of MAT items you receive)
4. Test scores are not unduly influenced by ancillary variables, such as writing ability
5. Success in the advanced course can be validly assessed
6. Test takers with high scores on the test will be more successful in the advanced course than test takers with low scores on the test

Adapted from 2014 Standards for Educational and Psychological Testing

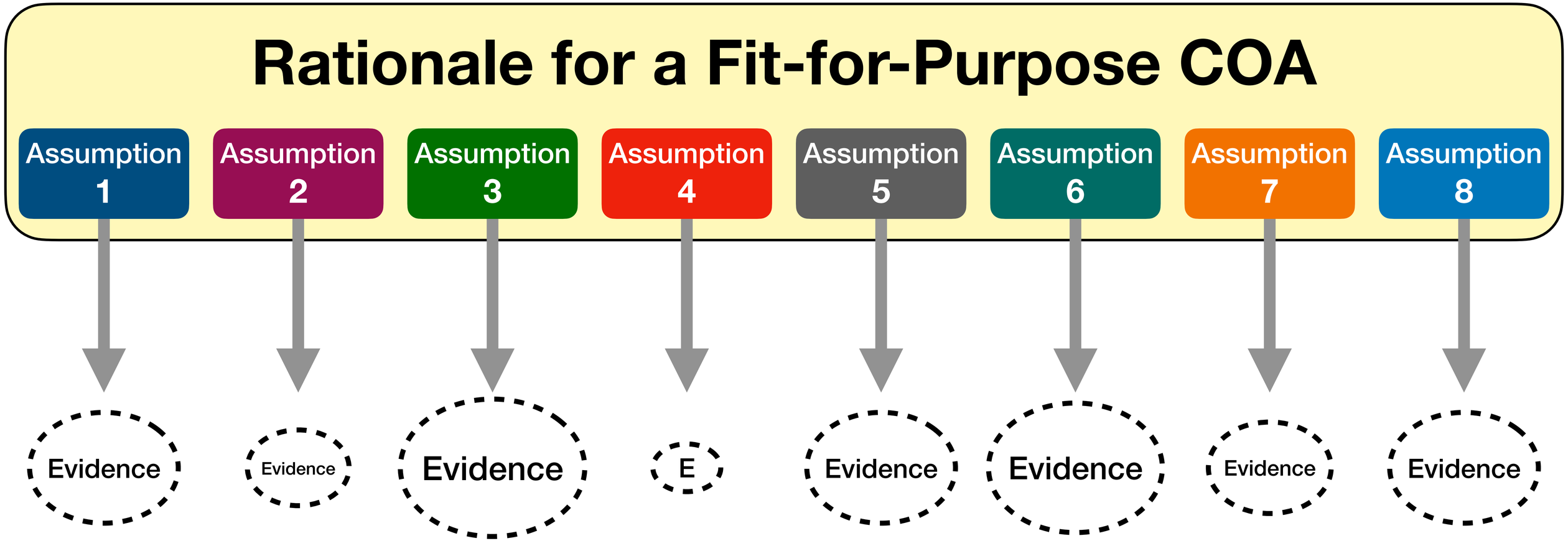
How do we know the type and amount of evidence needed in any given situation?

How do we know the type and amount of evidence needed in any given situation?

The amount of evidence needed for each assumption might differ depending on (1) a priori confidence that the assumption is true



The amount of evidence needed for each assumption might differ depending on (1) a priori confidence that the assumption is true and (2) the costs and benefits of collecting additional evidence in each case.



Some Benefits of Argument-Based Approach

Some Benefits of Argument-Based Approach

Focus on the rationale and how to support it, not on accumulating piles of information or “How much is good enough?”

Some Benefits of Argument-Based Approach

Focus on the rationale and how to support it, not on accumulating piles of information or “How much is good enough?”

Encourages consideration of the biggest threats to the rationale

Some Benefits of Argument-Based Approach

Focus on the rationale and how to support it, not on accumulating piles of information or “How much is good enough?”

Encourages consideration of the biggest threats to the rationale

Can be used for any type of measurement

Some Benefits of Argument-Based Approach

Focus on the rationale and how to support it, not on accumulating piles of information or “How much is good enough?”

Encourages consideration of the biggest threats to the rationale

Can be used for any type of measurement

Can be used for any proposed interpretation or use of scores

Qual Life Res (2018) 27:1711–1720
https://doi.org/10.1007/s11136-017-1644-z

SPECIAL SECTION: TEST CONSTRUCTION (BY INVITATION ONLY)

Fit for purpose and modern validity theory in clinical outcomes assessment

Michael C. Edwards[1,4] · **Ashley Slagle**[2] · **Jonathan D. Rubright**[3] · **R. J. Wirth**[4]

Accepted: 1 July 2017 / Published online: 7 July 2017

Abstract

Purpose The US Food and Drug Administration (FDA), as part of its regulatory mission, is charged with determining

Review After a brief introduction, the first section will review current ideas about "fit for purpose" and how it has been viewed by FDA. This section will also describe some

Quality of Life Research (2018) 27:1695–1710
https://doi.org/10.1007/s11136-018-1815-6

SPECIAL SECTION: TEST CONSTRUCTION (BY INVITATION ONLY)

Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity

Melanie Hawkins[1] · **Gerald R. Elsworth**[1] · **Richard H. Osborne**[1,2]

Accepted: 15 February 2018 / Published online: 20 February 2018

Abstract

Background Data from subjective patient-reported outcome measures (PROMs) are now being used in the health sector to make or support decisions about individuals, groups and populations. Contemporary validity theorists define validity not as a statistical property of the test but as the extent to which empirical evidence supports the interpretation of test scores for an intended use. However, validity testing theory and methodology are rarely evident in the PROM validation literature. Application of this theory and methodology would provide structure for comprehensive validation planning to support improved PROM development and sound arguments for the validity of PROM score interpretation and use in each new context.

Objective This paper proposes the application of contemporary validity theory and methodology to PROM validity testing.

Hawkins *et al. Journal of Patient-Reported Outcomes* (2021) 5:64
https://doi.org/10.1186/s41687-021-00332-y

Journal of Patient-Reported Outcomes

RESEARCH Open Access

Validity arguments for patient-reported outcomes: justifying the intended interpretation and use of data

Melanie Hawkins[1*], Gerald R. Elsworth[1], Sandra Nolte[2] and Richard H. Osborne[1]

Abstract

Background: Contrary to common usage in the health sciences, the term "valid" refers not to the properties of a measurement instrument but to the extent to which data-derived inferences are appropriate, meaningful, and useful for intended decision making. The aim of this study was to determine how validity testing theory (the *Standards for Educational and Psychological Testing*) and methodology (Kane's argument-based approach to validation) from education and psychology can be applied to validation practices for patient-reported outcomes

Quality of Life Research
https://doi.org/10.1007/s11136-021-02776-7

COMMENTARY

Constructing arguments for the interpretation and use of patient-reported outcome measures in research: an application of modern validity theory

Kevin P. Weinfurt[1]

Accepted: 18 January 2021

Abstract

The past 100 years have witnessed an evolution of the meaning of validity and validation within the fields of education and psychology. Validity was once viewed as a property of tests and scales, but is now viewed as the extent to which theory and evidence support proposed interpretations and uses of test scores. Uncertainty about what types of validity evidence were needed motivated the current "argument-based" approach, as reflected in the 2014 *Standards for Educational and Psychological Testing*. According to this approach, investigators should delineate the assumptions required in order for a proposed interpretation or use to be plausible and then seek evidence that supports or refutes those assumptions. Though

Table 2 Common assumptions that might comprise a rationale for the interpretation/use of patient-reported outcome measure scores for research purposes

A. The PROM's item content reflects all of the important aspects of the concept

B. Patients understand the items and response options as intended

C. Scores on the PROM are not unduly influenced by factors that are not part of the concept

1. *The PROM's item content does not include issues beyond the concept*
2. *Differences in linguistic/cultural backgrounds do not lead to substantially different interpretations of the items*
3. *Differences in patients' literacy or educational attainment do not lead to substantially different interpretations of the items*
4. *Errors of recollection do not unduly influence assessment of the concept* (for measures that use a recall period)
5. *Different modes of assessment do not lead to substantially different scores on the PROM*
6. *The patient's status on related, but separate, health domains does not unduly influence scores on the PROM*

D. The method of scoring responses to the item(s) of the PROM is appropriate for assessing the concept

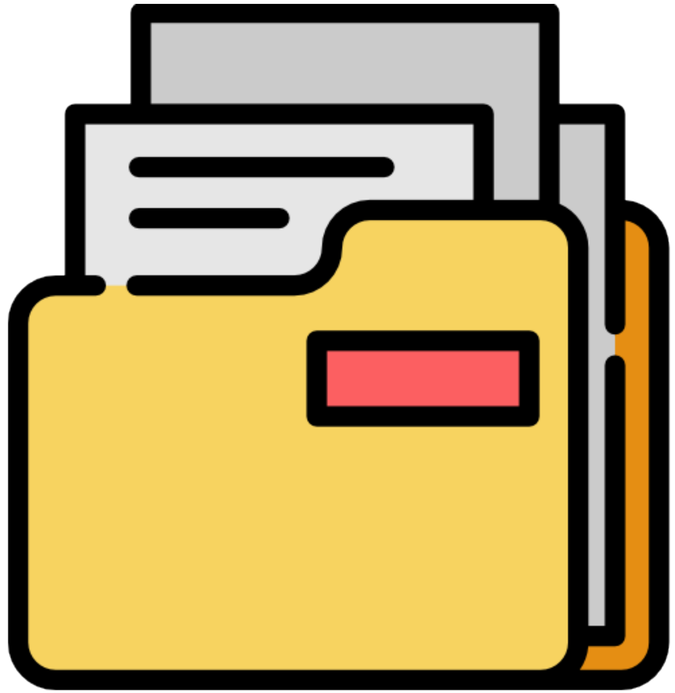
1. Scoring Inference
2. Scaling Inference
   a. The measurement model makes conceptual sense for the assessment of the concept and the items that are indicators of the concept
   b. In the case of a reflective or causal indicator model, the model provides acceptable fit to the response data
   c. Interpretation of scores is not unduly compromised by deviations from statistical assumptions of the model
   d. The scoring rule does not create bias with respect to one group of patients versus another

E. Scores from the PROM correspond to how patients actually feel and/or function in their daily lives

F. Scores from the PROM are sensitive enough to reflect differences in the concept between patients and/or within patients over time in levels of the concept being measured

Questions that need answering

- Would it be useful to construct a repository of validity arguments that have been made by sponsors?
  - Avoid redundant research
  - Educate people about what a successful validity argument looks like
- What is the most useful way to communicate validity arguments to regulators?
- How would a repository be managed? By whom?
- How can we strike a balance between transparency of information and competitive industry information that might be shared in a validity argument?
- Should the repository include validity arguments deemed to be unacceptable and acceptable by regulators?
- Track metrics – did the validity arguments ultimately result in more efficient preparation and review?
- How might the validity argument approach translate to regulatory bodies in other countries?



Questions and Contributions

Theresa Coles, PhD
Assistant Professor
Department of Population Health Sciences
Duke University School of Medicine

Theresa.Coles@Duke.edu

Kevin Weinfurt, PhD
Professor and Vice Chair of Research
Department of Population Health Sciences
Duke University School of Medicine
Kevin.Weinfurt@Duke.edu